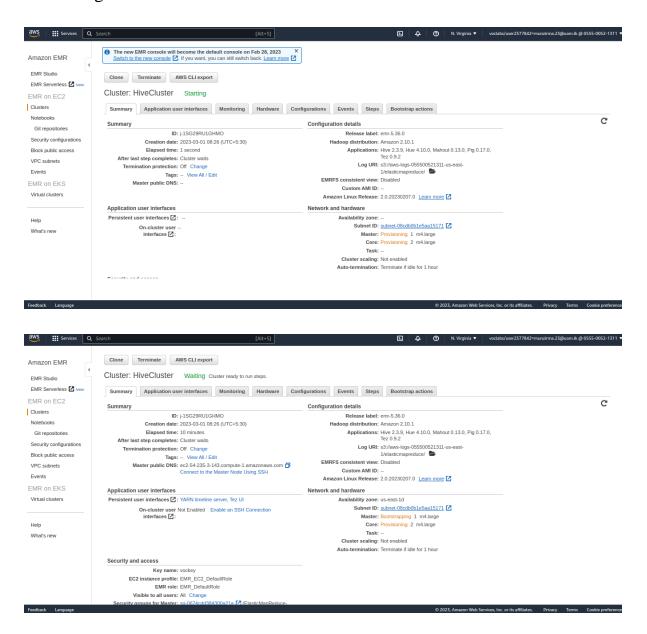
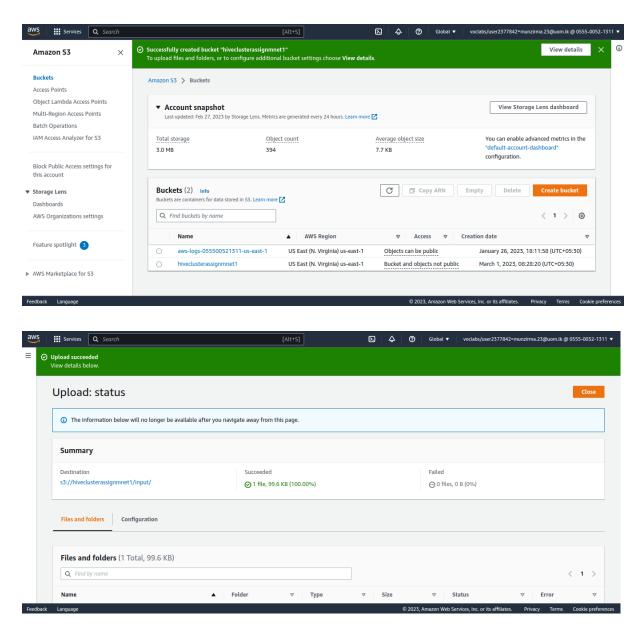
Hive

Screen Shots of the analysis

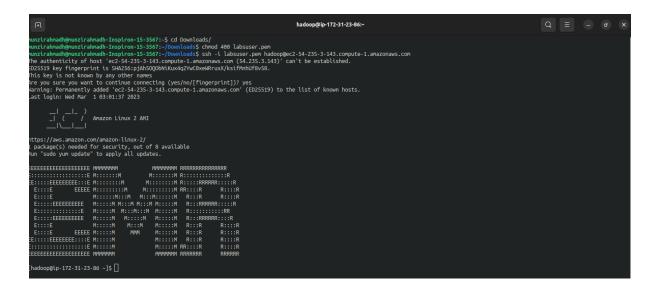
1. Launching an Amazon EMR cluster



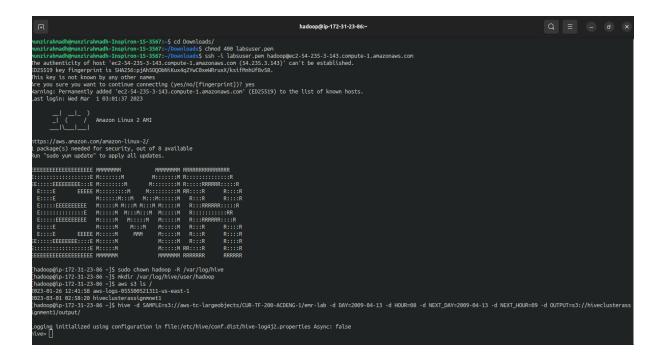
2. Create an S3 Bucket and Upload data and CSV file to S3 bucket.



3. Connecting to the Hadoop main node by using SSH



4. Running Hive interactively



5. Creating tables out of source data by using Hive and load data

6. Querying and analyzing the resulting dataset

```
at org.apache.hadoop.hive.cli.CliDriver.main(Clibriver.java:686)
at sun.reflect.NativeMethodAccessorImpl.invoke(Native Method)
at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)
at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodJava:49)
at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodJava:49)
at java.lang.reflect.Method.invoke(NethodJava:498)
at java.lang.reflect.MethodJava:498)
at java.lang.reflect.MethodJava:498)
at java.lang.reflect.MethodJava:498
at java.lang.reflect.MethodJava:498)
at java.lang.reflect.MethodJava:498
at java.lang.reflect.MethodJava:498
at java.lang.reflect.MethodJava:498
at java.lang.reflec
```

7. Other results

```
Use SELECT Year, avg((WeatherDelay /ArrDelay)*100) from delay_flights WHERE Year is NOT NULL GROUP BY Year

| Very 1D = hadoop_20230301032125_fb1ee526-5c01-49ea-bc0c-d3963ff08afd
| Very 1D = hadoop_20
```

```
| New Calenty - 7.297 Seconds, Petched: 8 POW(S)
| Inverse | Status | Petched: 8 POW(S)
| Inverse | Status | Petched: 8 POW(S)
| Inverse | Status | Petched: 8 POW(S)
| Inverse | Inverse
```

Year wise carrier delay from 2003-2010

Year wise NAS delay from 2003-2010

```
vevo SELECT Year, avg((MASDelay /ArrDelay)*190) from delay_flights MHERE Year is NOT NULL CROUP BY Year;
urry ID = hadoop_2023035162938_96f4d64d-2a74-40f4-b90e-edzc0fb9a3e9
stal jobs = 1
sunching Job 1 out of 1
atus: Running (Executing on YARN cluster with App id application_1678032641994_0002)
        _c1
_29.666276314267346
30.625925917941924
37.63693330628511
33.87551363464217
18.24570601769958
16.63868805373129
18.119312329937703
30.16552562594132
34en: 7.484 seconds, Fetched: 8 row(s)
ive> SELECT Year, avg((NASDelay /ArrDelay)*100) from delay.flights HHERE Year is NOT NULL CROUP BY Year;
uery ID = hadoon_20230305163044_8434d430-6c4a-4343-9f95-427d7a010ddb
otal jobs = 1
aunching Job 1 out of 1
tatus: Running (Executing on YARN cluster with App id application_1678032641994_0002)
  ive> SELECT Year, avg((NASDelay /ArrDelay)*100) from delay.flights WHERE Year is NOT NULL CROUP BY Year;
uery ID = hadoop_20230305163106_00368720-14a4-4e64-b414-2d2056f340dc
otal jobs = 1
aunching Job 1 out of 1
tatus: Running (Executing on YARN cluster with App id application_1678032641994_0002)
```

Year wise Weather delay from 2003-2010

```
ive> SELECT Year, avg((WeatherDelay /ArrDelay)*100) from delay_flights WHERE Year is NOT NULL GROUP BY Year;
uery ID = hadoop_20230305163258_0de24ce1-325e-4ed8-a1ea-000ca73a68ea
otal jobs = 1
aunching Job 1 out of 1
tatus: Running (Executing on YARN cluster with App id application_1678032641994_0002)
view- SELECT Year, avg((WeatherDelay /ArrDelay)*100) from delay_Flights WHERE Year is NOT NULL GROUP BY Year;
uery ID = hadoop_26239395163318_99b1b87-04cb-4ca1-b48c-7c862a3cb1d6

tati bis = 1
aunching Job 1 out of 1
tatus: Running (Executing on YARN cluster with App id application_1678832641994_0802)
          c1
7.831947964511205
4.042975783210287
0.45316615137982363
2.9923312955584664
6.447527997901555
5.85069715149616
4.588664183957953
3.77254499054088955
aken: 6.674 seconds, Fetched: 8 row(s)
ive> SELECT Year, avg((MeatherDelay /ArrDelay)*100) from delay_flights WHERE Year is NOT NULL GROUP BY Year;
uery ID = hadoop_20230305163344_183b7dd2-e4f2-4f76-9374-4eec107248a6
otal jobs = 1
aunching Job 1 out of 1
tatus: Running (Executing on YARN cluster with App id application_1678032641994_0002)
           _c1
7.8319479664511205
4.047975783210287
0.45316615137982363
2.9023312955584664
6.4475279976916555
5.85069715149616
4.588664183957953
3.77254499054088955
aken: 7.865 seconds, Fetched: 8 row(s)
use useem. 1.220 seconds, recurso, o rom(s)
ive> SELECT Year, avg((WeatherDelay /ArrDelay)*100) from delay_flights WHERE Year is NOT NULL GROUP BY Year;
uery ID = hadoop_2028035163439_cc0aa1b5-32e4-4be8-9481-282a42d85199
otal jobs = 1
aunching Job 1 out of 1
tatus: Running (Executing on YARN cluster with App id application_1678032641994_0002)
```

```
| New | Select | Very | Average | Av
```

```
| Number | N
```

Year wise security delay from 2003-2010

8. Terminating the Cluster and cleaning the S3 bucket

